JOURNAL OF MULTIVARIATE ANALYSIS 7, 572–583 (1977)

Multivariate Nonparametric Tests for Independence

BIMAL KUMAR SINHA* AND H. S. WIEAND†

University of Pittsburgh

Communicated by P. K. Sen

Multivariate generalizations of Bhuchongkul's bivariate rank statistics [*Ann. Math. Statist.* **35** (1964)] have been introduced and studied in this paper for the purpose of testing multivariate independence. It is shown that the test statistics can be expressed as rank statistics which are easy to compute, have asymptotic normal distributions, and can detect mutual dependence in alternatives which are pairwise independent. The tests are compared to the Puri–Sen–Gokhale [*Sankyhā Ser. A* **32** (1970)] tests and a normal theory test [Anderson, "An Introduction to Statistical Analysis," Wiley, 1958] using Pitman efficiency.

1. INTRODUCTION

Nonparametric tests of bivariate independence have been discussed by statisticians for more than 50 years and numerous nonparametric statistics for testing bivariate independence are available. However, very little has been written regarding nonparametric tests of independence involving three or more variates. Two notable exceptions are papers by Blum *et al.* [4] and Puri *et al.* [8]. The former paper introduced the idea of using functionals of the empirical distribution function to detect dependence. In particular, two statistics—one a generalization of a Kolmogorov–Smirnov statistic and the other a generalization of a Cramer–von Mises statistic— were introduced. The statistics appear to have excellent power properties, but are not easily computed, and do not have known asymptotic distributions under the hypothesis, hence at the present time are not very practical. The statistics proposed by Puri *et al.* are generalizations of linear rank statistics. These statistics have known asymptotic distributions, are fairly easy to compute, and in fact are useful for many problems.

Received May 17, 1977; revised July 12, 1977.

AMS 1970 subject classifications: Primary 62G10; Secondary 62F20.

Key words and phrases: Tests of independence, empirical distribution function, ranks, consistency, Bahadur and Pitman efficiencies.

* On leave from the Indian Statistical Institute, Calcutta, India.

† Research supported in part by a grant from the Faculty of Arts and Sciences, University of Pittsburgh.

ISSN 0047-259X

However, the statistics are essentially designed to test pairwise independence and will not detect mutual dependence if the variates are pairwise independent.

Our object in this paper is to introduce multivariate tests, using generalizations of Bhuchongkul's rank statistics, which are easy to compute, are designed to detect mutual dependence, and are asymptotically normal under both the hypothesis and alternative. In Section 2 the test statistics are defined and their asymptotic normality is proved. In Section 3 we compare our test statistics to other known test statistics using Pitman efficiency and discuss some advantages and disadvantages of our tests. This section also includes examples of alternatives which have mutual dependence but pairwise independence for which our tests are consistent but the Puri–Sen–Gokhale tests and the standard normal theory test are not.

Remark. In a paper which appeared while this paper was being revised, Simon [11] considered a multivariate generalization of Kendall's tau statistic which is also nonparametric. He showed how the statistic could be used for data reduction. It does not appear that this statistic is applicable for the problem we are considering, but a careful reading of his paper may show otherwise.

2. The Test Statistics and Their Asymptotic Normality

Let $\mathbf{X}_j' = (X_{1j}, \ldots, X_{pj})$, $j = 1, \ldots, n$ be n i.i.d. r.v.'s having a p-variate continuous c.d.f. $F(\mathbf{x})$, $\mathbf{x} \in E^p$, the Euclidean p-space. Let the marginal c.d.f.'s of $X_1, \ldots, X_p$ be denoted by $F_1(x_1), \ldots, F_p(x_p)$, $x_i \in E$, $i = 1, \ldots, p$. We consider the problem of testing the null hypothesis that the variables $X_1, \ldots, X_p$ are mutually independent, i.e.,

$$H_0\colon F(\mathbf{x}) = \prod_{i=1}^{p} F_i(x_i), \qquad \text{for all } \mathbf{x} \in E^p. \tag{2.1}$$

Define the empirical marginal c.d.f.'s based on the $\mathbf{x}_j$'s

$$F_{in}(x_i) = \frac{1}{n} \sum_{j=1}^{n} I_{(-\infty, x_i]}(x_{ij}), \qquad x_i \in E, \quad i = 1, \ldots, p \tag{2.2}$$

and also the empirical joint c.d.f.

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{p} I_{(-\infty, x_i]}(x_{ij}), \qquad \mathbf{x} \in E^p. \tag{2.3}$$

Note that for $p = 2$ Bhuchongkul's statistic T_N is given by (Bhuchongkul [3])

$$S_n^{(2)}(\mathbf{J}_n) = n^{1/2} \iint_{E^2} J_{1n}(F_{1n}(x_1))\, J_{2n}(F_{2n}(x_2))\, dF_n(x_1, x_2),$$

where the J_{in}'s are suitably defined score functions.

As a generalization of this, we propose the following statistics to test the null hypothesis H_0:

$$S_n^{(p)}(\mathbf{J}_n) = n^{1/2}\left\{\int\cdots\int_{E^p} J_{1n}(F_{1n}(x_1))\cdots J_{pn}(F_{pn}(x_p))\,dF_n(\mathbf{x})\right\} = n^{-(p+1/2)}\sum_{j=1}^{n}\prod_{i=1}^{p} J_{in}(R_{ij}), \tag{2.4}$$

large absolute values of $S_n^{(p)}(\mathbf{J}_n)$ being significant, where R_{ij} is the rank of the jth observation among the n values of the ith variate, $j = 1,\ldots, n$, $i = 1,\ldots, p$, and the J_{in}'s are score functions satisfying some conditions stated below.

Asymptotic normality of $S_n^{(p)}(\mathbf{J}_n)$. Define $I_{in} = \{x: 0 < F_{in}(x) < 1\}$, $i = 1,\ldots, p$.

THEOREM 1. *If*

(1) $J_i(u) = \lim_{n\to\infty} J_{in}(i)$, $i = 1,\ldots, p$ *exist for* $0 < u < 1$ *and are not constants,*

(2) $\int_{I_{1n}\times\cdots I_{pn}} (\prod_{i=1}^{p} J_{in}(F_{in}(x_i)) - \prod_{i=1}^{p} J_i(F_{in}(x_i))\,dF_n(x) = o_p(n^{-1/2})$,

(3) $J_{in}(1) = o(n^{(1/2p)})$, $i = 1,\ldots, p$,

(4) $|J_i(u)| \leqslant K[u(1-u)]^{-\alpha}$ *for some* $0 < \alpha < 1/2p$, $i = 1,\ldots, p$,
$|J_i'(u)| \leqslant K[u(1-u)]^{-1}$, $|J_i''(u)| \leqslant K[u(1-u)]^{-2}$, $i = 1,\ldots, p$,

then

$$\lim_{n\to\infty} P\left\{\frac{S_n^{(p)}(\mathbf{J}_n) - \mu}{\sigma_p} \leqslant t\right\} = (2\pi)^{-1/2}\int_{-\infty}^{t} e^{-x^2/2}\,dx$$

uniformly with respect to F_i's *and* F, *provided* $\sigma_p \neq 0$; *where*

$$\mu = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{i=1}^{p} J_i(F_i(x_i))\,dF(\mathbf{x}) \tag{$*$}$$

and

$$\sigma_p^2 = \operatorname{Var}\left\{\prod_1^p J_i(F_i(x_i)) + \sum_{i=1}^{p}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\{\phi_{x_i}(u) - F_i(u)\}\,J_i'(F_i(x))\prod_{j\neq i} J_j(F_j(x_j))\,dF(\mathbf{x})\right\}, \tag{$**$}$$

where $\phi_{x_i}(u) = 1$ *if* $x_i \leqslant u$ *and is zero otherwise.*

Proof. The proof is along the same lines as in Bhuchongkul [3]. $S_n^{(p)}(\mathbf{J}_n)$ can be written as

$$S_n^{(p)}(\mathbf{J}_n) = n^{1/2}\int\cdots\int\prod_{i=1}^{p} J_{in}(F_{in}(x_i))\,dF_n(\mathbf{x}).$$

Now

$$\prod_{i=1}^{p} J_{in}(F_{in}(x_i)) = \prod_{i=1}^{p} J_{in}(F_{in}(x_i)) - \prod_{i=1}^{p} J_i(F_{in}(x_i)) + \prod_{i=1}^{p} J_i(F_{in}(x_i))$$

and using Taylor's expansion, we can express $\prod_{i=1}^{p} J_i(F_{in}(x_i))$ as

$$\begin{aligned}
&\prod_{i=1}^{p} J_i(F_{in}(x_i)) \\
&\quad = \prod_{i=1}^{p} J_i(F_i(x_i)) + \sum_{i=1}^{p} (F_{in}(x_i) - F_i(x_i))\, J_i'(F_i(x_i)) \prod_{j\neq i} J_j(F_j(x_j)) \\
&\qquad + \tfrac{1}{2} \sum_{i} (F_{in}(x_i) - F_i(x_i))^2\, J_i''[\theta F_{in} + (1-\theta)F_i] \prod_{j\neq i} J_j[\theta F_{jn} + (1-\theta)F_j] \\
&\qquad + \sum_{i<j}\sum (F_{in}(x_i) - F_i(x_i))(F_{jn}(x_j) - F_j(x_j))\, J_i'(\theta F_{in} + (1-\theta)F_i) \\
&\qquad \times J_j'(\theta F_{jn} + (1-\theta)F_j) \prod_{K\neq i,j} J_K(\theta F_{Kn} + (1-\theta)F_K).
\end{aligned}$$

Define

$$A_{on} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} J_i(F_i(x_i))\, dF_n(\mathbf{x}),$$

$$A_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (F_{in}(x_i) - F_i(x_i))\, J_i'(F_i(x_i)) \prod_{j\neq i} J_j(F_j(x_j))\, dF(\mathbf{x}), \qquad i = 1,\dots, p,$$

$$B_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (F_{in}(x_i) - F_i(x_i))\, J_i'(F_i(x_i)) \prod_{j\neq i} J_j(F_j(x_j))\, d(F_n(\mathbf{x}) - F(\mathbf{x})), \qquad i = 1,\dots, p,$$

$$\begin{aligned}
C_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} &(F_{in}(x_i) - F_i(x_i))^2 \\
&\times J_i''(\theta F_{in} + (1-\theta)F_i) \prod_{j\neq i} J_j(\theta F_{jn} + (1-\theta)F_j)\, dF_n(\mathbf{x}), \qquad i = 1,\dots, p,
\end{aligned}$$

$$\begin{aligned}
D_{ijn} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} &(F_{in}(x_i) - F_i(x_i))(F_{jn}(x_j) - F_j(x_j))\, J_i'(\theta F_{in} + (1-\theta)F_i) \\
&\times J_j'(\theta F_{jn} + (1-\theta)F_j) \prod_{K\neq i,j} J_K(\theta F_{Kn} + (1-\theta)F_K)\, dF_n(\mathbf{x}), \\
&\qquad i < j, \quad i, j = 1,\dots, p,
\end{aligned}$$

$$B_n^* = \int \cdots \int_{I_{1n}\times\cdots\times I_{pn}} \left\{ \prod_{i=1}^{p} J_{in}(F_{in}(x_i)) - \prod_{i=1}^{p} J_i(F_{in}(x_i)) \right\} dF_n(\mathbf{x}),$$

$$C_n^* = \int \cdots \int_{R^p - I_{1n}\times\cdots\times I_{pn}} \prod_{i=1}^{p} J_{in}(F_{in}(x_i))\, dF_n(\mathbf{x}),$$

and

$$D_n^* = -\int\cdots\int_{R^p - I_{1n}\times\cdots\times I_{pn}} \left[\prod_{i=1}^{p} J_i(F_i(x_i)) + \sum_i (F_{in}(x_i) - F_i(x_i)) \right.$$

$$\left. \times\, J_i'(\theta F_{in} + (1-\theta)F_i) \prod_{j\neq i} J_j(\theta F_{jn} + (1-\theta)F_j)\right] dF_n(\mathbf{x}).$$

We will show that $\sum_{i=0}^{p} A_{in}$ with a suitable normalization has a limiting normal distribution and that the rest of the terms are all $o_p(n^{-1/2})$. The arguments are exactly similar to those of Bhuchongkul [3] and we therefore omit the details. The asymptotic negligibility

(i) of C_n^* and B_n^* follows immediately from assumptions (1), (2), and (3) of Theorem 1;

(ii) of D_n^* follows from that of B_{6n} in Bhuchongkul [3];

(iii) of C_{in}'s and D_{ijn}'s follows from those of B_{3n}, B_{4n}, and B_{5n} in Bhuchongkul [3];

(iv) of B_{in}'s follows from those of B_{1n} and B_{2n} of Bhuchongkul [3].

Finally, it can be shown easily as in Bhuchongkul [3] that $\sum_{i=0}^{p} A_{in}$ is the average of n i.i.d. variables with mean μ given in (*), variance σ_p^2 given in (**), and finite third moment. Hence the theorem. Q.E.D.

Remarks. (a) The first inequality of assumption (4) is more restrictive than the corresponding one imposed by Bhuchongkul [3] or Chernoff and Savage [5]. However, considering $J = F^{-1}$, (4) is satisfied by several distribution functions F such as normal, exponential, logistic, and uniform. In all these cases, if $J_{in}(l/n)$ is the expectation of the lth-order statistic of samples of size n from a population with c.d.f. $J_i^{-1} = F_i$, assumptions (1), (2), and (3) are also satisfied. This can be easily checked using the results of Theorem 2 of Chernoff and Savage [5] as in Theorem 2 of Bhuchongkul [3].

(b) A referee noted that Ruymgaart *et al.* [10] were able to eliminate the condition involving the second derivative in assumption (4) in the two-dimensional case. It seems likely that their techniques would permit us to eliminate this condition also, but we have not verified this.

(c) Under the hypothesis H_0 of independence,

$$\mu = \prod_{i=1}^{p} \int_{-\infty}^{\infty} J_i(u)\, du$$

and

$$\sigma_p^2 = \mathrm{Var}_{H_0}\left[\prod_{i=1}^{p} J_i(F_i(X_i))\right] - \sum_{i=1}^{p} \prod_{j\neq i} E_{H_0}^2(J_j(F_j(X_j)))\, \mathrm{Var}_{H_0}(J_i(F_i(X_i))).$$

In particular, taking the J_i's as identity, Spearman's version of $S_n^{(p)}(\mathbf{J}_n)$ has under H_0 mean 2^{-p} and variance $\sigma_p^2 = [3^{-p} - 4^{-p}(p+3)/3]$ while taking $J = \Phi^{-1}$, the normal scores version of the statistic $S_n^{(p)}(\mathbf{J}_n)$ has under H_0 mean 0 and variance $\sigma_p^2 = 1$.

3. Efficiency

In this section we will consider several types of dependence and compare the ability of $S_{(p)}^n$ to detect the dependence with appropriate competitors. Most of our computations will be done using Spearman's version of $S_n^{(p)}(\mathbf{J}_n)$, in which case we will denote the statistic by $S_n^{(p)}$, i.e.,

$$S_n^{(p)} = n^{1/2} \int \cdots \int \prod_{i=1}^{p} F_{in}(x_i)\, dF_n(x)$$

$$= n^{-(p+1)/2} \sum_{j=1}^{n} \prod_{i=1}^{p} R_{ij}\,.$$

For nonparametric competitors we will consider the Puri–Sen–Gokhale [8] statistics which have the form

$$V_n^J(p) = (n-1) \sum_{i<j=1}^{p} T_{n,ij}^2 / T_{n,ii} T_{n,jj} \tag{3.1}$$

where

$$T_{n,ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_{in}(nF_{in}(x_i)/(n+1))\, J_{jn}(nF_{jn}(x_j)/(n+1))\, dF_{ijn}(x_i\,, x_j).$$

F_{ijn} denotes the two-dimensional empirical distribution function over the ith and jth variates and J_n's are score functions. For the score functions considered by Puri *et al.* [8], $V_{n(p)}^J/n^{1/2}$ converges in probability to $\sum_{i<j}^{p} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_i(F_i(x_i))\, J_j(F_j(x_j))\, dF_{ij}(x_i\,, x_j)$ where F_{ij} denotes the two-dimensional distribution function of $(X_i\,, X_j)$ and the J's satisfy $J_i = \lim_n J_n^{(i)}$. If the X_i's are pairwise independent $V_{n(p)}^J$ is asymptotically $\chi^2(p(p-1)/2)$. A parametric competitor will be a statistic defined in Anderson [1]

$$U_n^{(p)} = |\,\mathbf{A}\,| \Big/ \prod_{i=1}^{p} a_{ii}\,, \tag{3.2}$$

where $\mathbf{A} = ((a_{ij})) = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$, $\bar{\mathbf{X}} = \sum_{i=1}^{n} \mathbf{X}_i/n$. Under assumptions which are satisfied in the examples discussed below, pairwise independence of the X_i's implies $-n \log U_n^{(p)}$ is asymptotically $\chi^2(p(p-1)/2)$.

For our first example, we will let $\mathbf{X}$ have the density

$$f(\mathbf{x}) = (2\pi)^{-p/2} \exp\left\{-\sum_{i=1}^{p} x_i^2/2\right\}\left(1 + \left(\prod_{i=1}^{p} x_i \exp\{-(x_i^2 - 1)/2\}\right)\right)$$
$$\text{for} \quad -\infty < x_1, \ldots, x_p < \infty. \tag{3.3}$$

With this density, we have exchangeable random variables which have $N(0, 1)$ marginals, l-variate $N(0, I_{l\times l})$ distributions for $l < p$, and yet the variates are mutually dependent. Due to the pairwise independence of the variates, the Puri–Sen–Gokhale statistics and $U_n^{(p)}$ would not form consistent tests for the hypothesis of independence against this alternative, however, $| S_n^{(p)} - 2^{-p} |/n^{1/2}$ converges in probability to $[4^{-1}(2e/3\pi)^{1/2}]^p$, hence would form a consistent test which means it would be "infinitely" more efficient than $U_n^{(p)}$ or any $V^J_{(n)p}$.

Of course, similar examples can be constructed with nonnormal marginals, e.g.,

$$f(x) = \left(1 + \prod_{i=1}^{p} x_i\right)\Big/2^p, \qquad -1 < x_1, \ldots, x_p < 1, \tag{3.4}$$

in which case $| S_n^{(p)} - 2^{-p} |/n^{1/2}$ converges in probability to 6^{-p} while the Puri–Sen–Gokhale statistics and $U_n^{(p)}$ would again fail to form consistent tests. The significant fact in these examples is that $S_n^{(p)}$ does admit the possibility of detecting mutual dependence, despite pairwise independence, while the Puri–Sen–Gokhale statistics and $U_n^{(p)}$ do not.

The next natural question is how does $S_n^{(p)}$ compare to $V^J_{(n)p}$ and $U_n^{(p)}$ when there is pairwise dependence. In order to address this question, we have chosen the following model.

Let $U_1, \ldots, U_p$ be independent random variables with densities (distribution functions) $f_1, \ldots, f_p$ $(F_1, \ldots, F_p)$, respectively. Assume f_i', f_i'', and f_i''' are bounded and continuous for each i and that each f_i is strictly positive on $(-\infty, \infty)$. Let W be a random variable, independent of the U_i's, with finite second moment. Define $X_1, \ldots, X_p$ by $X_i = U_i + \theta k_i W$, $i = 1, \ldots, p$, where the k_i's are real constants and $0 < \theta < \infty$. This alternative is an extension of a model in Bhuchongkul [3]. We will compute the limiting (as $\alpha \to 0$) Pitman efficiencies of $S_n^{(p)}$ relative to $V^J_{(n)p}$ and $U_n^{(p)}$ for alternatives taken from this model.

The usual tools for computing the Pitman efficiencies (Fraser [6] and Puri and Sen [9]) will not work in this case because the asymptotic distribution of $S_n^{(p)}$ does not have the same form as that of $V^J_{(n)p}$ or $U_n^{(p)}$. However, it is possible to compute the limiting (as $\alpha \to 0$) Pitman efficiency using the approximate Bahadur efficiency. The method is to compute the approximate Bahadur slopes of two statistics and verify that the two statistics satisfy an additional condition, III* (Wieand [12]). If so, the limit (as the alternative approaches the hypothesis) of the approximate Bahadur efficiency of the two statistics represents the limit (as $\alpha \to 0$) of their exact Pitman efficiency.

To begin, consider the statistic $| S_n^{(p)*} | = | S_n^{(p)} - 2^{-p} |$ which is equivalent to $| S_n^{(p)} |$. Under the hypothesis $S_n^{(p)*}$ is asymptotically $N(0, \sigma_p^2)$, where $\sigma_p^2 = 3^{-p} - 4^{-p}(p+3)/3$, and under the alternative $| S_n^{(p)*} |/n^{1/2}$ converges in probability to $| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_i(x_i)\, dF(x_1 ,..., x_p) - 2^{-p} |$. For the alternatives given above, it follows that in Bahadur's notation (Bahadur [2]) we have

$$a = \sigma_p^{-2} \tag{3.5}$$

and

$$b(\theta) = \left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(x_1 ,..., x_p) - 2^{-p} \right|, \tag{3.6}$$

where $F_{\theta i}$ is the distribution function of X_i and $F_\theta(x_1 ,..., x_p)$ is the distribution of $\mathbf{X}$ under the alternative. Using Taylor expansions, we find

$$b(\theta) = \theta^2 \operatorname{var}(W) 2^{2-p} \left(\sum_{i<j=1}^{p} k_i k_j \left(\int f_i^2(x_i)\, dx_i \right) \left(\int f_j^2(x_j)\, dx_j \right) \right) + o(\theta^2), \tag{3.7}$$

where f_i represents the density of U_i. Finally, the approximate Bahadur slope is

$$c_S(\theta) = \sigma_p^{-2} 4^{2-p} \theta^4 \operatorname{var}^2(W) \left(\sum_{i<j=1}^{p} k_i k_j \left(\int f_i^2(x_i)\, dx_i \right) \left(\int f_j^2(x_j)\, dx_j \right) \right)^2 + o(\theta^4). \tag{3.8}$$

To verify that $| S_n^{(p)*} |$ satisfies condition III*, it must be shown that there is a $\theta' > 0$ such that for every $\epsilon > 0$ and $\delta \in (0, 1)$, there is an n' such that $n > n'/b^2(\theta)$ and $\theta \in (0, \theta')$ implies $P_\theta\{| \, | S_n^{(p)*} |/n^{1/2} - b(\theta)| < \epsilon b(\theta)\} > 1 - \delta$. It follows from (3.7) that we may choose a θ' such that $\theta \in (0, \theta')$ implies $b(\theta) < 1$. From the proof of asymptotic normality of $S_n^{(p)}(\mathbf{J}_n)$ given in Section 2, we have

$$\begin{aligned}
&| \, | S_n^{(p)*} |/n^{1/2} - b(\theta)| \\
&\quad \leqslant |(| A_{on} - 2^{-p} | - b(\theta))| + \sum_{i=1}^{p} | A_{in} | + \sum_{i=1}^{p} | B_{in} | + | R_n | \\
&\quad \leqslant \left| A_{on} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(\mathbf{x}) \right| + \sum_{i=1}^{p} | A_{in} | + \sum_{i=1}^{p} | B_{in} | + | R_n | \\
&\quad \leqslant \left| A_{on} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(\mathbf{x}) \right| + \sum_{i=1}^{p} | A_{in} | \\
&\qquad + \sum_{i=1}^{p} | B_{in} | + \sum_{l=2}^{p} \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant p} \sup_{x_{i_1}} | F_{i_1 n}(x_{i_1}) - F_{i_1}(x_{i_1})|
\end{aligned} \tag{3.9}$$

and this last expression can be written as $| A_{0n} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(\mathbf{x})|$ and a finite number of terms each of which is less than or equal to $\sup_x | F_{in}(x) -$

$F_i(x)|$ for some i. It is shown in Wieand [12] that for each i and any $\epsilon' > 0$ and $\delta' \in (0, 1)$ there is an n_i^* such that $n > n_i^*/b^2(\theta)$ and $\theta \in (0, \theta')$ imply $P(\sup_{x_1} |F_{in}(x_i) - F_i(x_i)| < \epsilon' b(\theta)) > 1 - \delta$. It follows that with the proper choice of ϵ' and δ' and $n_1 = \max_i n_i^*$, $n > n_1/b^2(\theta)$ and $\theta \in (0, \theta')$ would imply

$$P\left(\sum_{i=1}^{p} | A_{in} | + \sum_{i=1}^{p} | B_{in} | + | R_n | \leqslant (\epsilon/2)\, b(\theta)\right) > 1 - \delta/2. \tag{3.10}$$

Finally, the Berry–Esseen bound assures us that

$$\left| P\left(n^{1/2}\left(A_{on} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(x_1, ..., x_p)\right) \sigma_A^{-1} < z\right) - \Phi(z) \right|$$
$$\leqslant k \cdot n^{-1/2} \tag{3.11}$$

for all z, where $\sigma_A{}^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}^2\, dF_\theta(x_1, ..., x_p)$. We can choose M such that $\Phi(M^{1/2}\epsilon/2) > 1 - \delta/8$ and $n_2 > M$ such that $kn_2^{-1/2} < \delta/8$. Then $n > n_2/b_2(\theta)$ and $\theta \in (0, \theta')$ implies

$$P\left(n^{1/2} \left| A_{on} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(x_1, ..., x_p) \right| \sigma_A^{-1} < n^{1/2} b(\theta)\epsilon/2\right) > 1 - \delta/2$$

which implies

$$P\left\{\left| A_{on} - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} F_{\theta i}(x_i)\, dF_\theta(x_1, ..., x_p) \right| < \epsilon b(\theta)/2\right\} > 1 - \delta/2 \tag{3.12}$$

since $\sigma_A{}^2 \leqslant 1$. Letting $n' = \max(n_1, n_2)$, it follows from (3.9), (3.10), and (3.11) that $n > n'/b^2(\theta)$ and $\theta \in (0, \theta')$ implies $P\{|(| S_n^{(p)*} |/n^{1/2} - b(\theta))| < \epsilon b(\theta)\} > 1 - \delta$, hence condition III*. Similarly, the Puri–Sen–Gokhale statistic with the score function

$$J_n^{(1)}(i/(n + 1)) = (12/(n^2 - 1))^{1/2}(i - (n + 1)/2) \tag{3.13}$$

(the Spearman score function) satisfies condition III* with

$$c_V(\theta) = 144\theta^4 \operatorname{var}^2(W) \sum_{i<j=1}^{p} k_i^2 k_j^2 \left(\int f_i^2(x_i)\, dx_i\right)^2 \left(\int f_j^2(x_j)\, dx_j\right)^2 + o(\theta). \tag{3.14}$$

The proof of this fact is similar to that given above and is omitted.

Letting $E_{S_n^{(p)} V_n^{J_1(p)}}$ represent the limiting (as $\alpha \to 0$) Pitman efficiency, we have, by the theorem given in Wieand [12] $E_{S_n^{(p)} V_n^{J_1(p)}} = \lim_{\theta \to 0}(c_S(\theta)/c_V(\theta))$, i.e., by (3.8) and (3.14)

$$E_{S_n^{(p)} V_n^{J_1}{}_{(p)}} = \frac{1}{144\sigma_p{}^2 4^{p-2}} \frac{\{\sum_{i<j=1}^{p} k_i k_j (\int f_i^2(x_i)\, dx_i)(\int f_j^2(x_j)\, dx_j)\}^2}{\sum_{i<j=1}^{p} k_i^2 k_j^2 (\int f_i^2(x_i)\, dx_i)^2 (\int f_j^2(x_j)\, dx_j)^2}. \tag{3.15}$$

It is immediately obvious that if we admit negative k_i's, which would represent some negative correlation, $E_{S_n^{(p)} V_n^{J_1}(p)}$ can be 0, i.e., $S_n^{(p)}$ can be "infinitely" less efficient than $V_n^{J_1}(p)$.

It is again possible to consider the case of exchangeable random variables with normal marginals with the above model by letting the U_i's and W be $N(0, 1)$ and letting each $k_i = p^{-1/2}$. For competitors of $S_n^{(p)}$, we will use two Puri–Sen–Gokhale statistics, one with score function $J_n^{(1)}$ given by (3.13) and the other with score function $J_n^{(2)}(i/(n+1)) = E(Z_n^{(i)})$ (the normal score function) where $Z_n^{(i)}$ is the ith order statistic (out of n) from a $N(0, 1)$ population. We will also use $U_n^{(p)}$ and the likelihood ratio statistic for this alternative,

$$L_n^{(p)} = (p/n^{1/2}) \sum_{i=1}^{n} \bar{x}_i^2/2^{1/2} - (n/2)^{1/2}, \tag{3.16}$$

where $\bar{x}_i = \sum_{j=1}^{p} x_{ji}/p$, $i = 1,\dots, n$.

Our reason for including $E_{S_n^{(p)} L_n^{(p)}}$ is that $L_n^{(p)}$ is essentially the "optimal" statistic for this alternative.

Remark. We know the limiting Pitman efficiency $E_{S_n^{(p)} V_n^{J_1}(p)}$ for alternatives of this type from (3.15). The Pitman efficiency of $V_n^{J_1}(p)$ to $V_n^{J_2}(p)$ and $V_n^{J_2}(p)$ to $U_n^{(p)}$ are given by Puri and Sen [9] (and since these Pitman efficiencies are independent of α, these are the limiting (as $\alpha \to 0$) Pitman efficiencies as well). Noting that, in general, $E_{S_n T_n} = E_{S_n V_n} E_{V_n T_n}$, this permits us to compute $E_{S_n^{(p)} V_n^{J_2}(p)}$ and $E_{S_n^{(p)} U_n^{(p)}}$. Finally, $L_n^{(p)}$ satisfies condition III* with $c_L(\theta) = \theta^4/2$ which can be shown using a technique similar to that given for $S_n^{(p)}$.

In Table I, we give the efficiencies of $S_n^{(p)}$ to the alternatives mentioned above for $p = 2(1)10$. The efficiency is given for general p as well.

TABLE I

Limiting Pitman Efficiencies for Normal Alternatives

p	$E_{S_n^{(p)} V_n^{J^1}(p)}$	$E_{S_n^{(p)}, U_n^{(p)}}$	$E_{S_n^{(p)} L_n^{(p)}}$
2	1	0.9119	0.4559
3	0.9	0.8207	0.5471
4	0.8059	0.7349	0.5511
5	0.7181	0.6548	0.5238
6	0.6365	0.5804	0.4837
7	0.5611	0.5117	0.4386
8	0.4921	0.4487	0.3926
9	0.4292	0.3914	0.3479
10	0.3725	0.3397	0.3057
$\sim$	$\dfrac{p(p-1)}{18\sigma_p^2 4^p}$	$\dfrac{p(p-1)}{2\pi^2\sigma_p^2 4^p}$	$\dfrac{(p-1)^2}{2\pi^2 4^p \sigma_p^2}$

For this particular example, $E_{S_n^{(p)} V_n^{J_2}(p)} = E_{S_n^{(p)} U_n^{(p)}}$ so only $E_{S_n^{(p)} U_n^{(p)}}$ is included in the table.

It is possible to compute the efficiency of $S_n^{(p)}$ to $S_n^{(p)}(\mathbf{J}_n)$ for any of the score functions given in Section 2 using the standard Pitman technique. However, for the alternatives considered above, $E_{S_n^{(p)} S_n^{(p)}(\mathbf{J}_n)} = \infty$ for $p > 2$ if $\int J(u)\, du = 0$. We believe that the only significance of this fact is that our alternative model is inappropriate for such score functions. Of course, other types of dependence arising from different models could be considered. However, any model designed for computing efficiencies easily is probably somewhat unrealistic so we did not attempt this comparison. We believe the results already obtained are sufficient to justify the following conclusions. If the hypothesis of independence is to be tested against pairwise dependence, the Puri–Sen–Gokhale tests and/or the normal theory tests are better than our proposed tests. However, if the alternative is such that there may be mutual dependence without pairwise dependence, one of our tests should be used (possibly in conjunction with $V_n^J(p)$ or $U_n^{(p)}$), since, to the best of our knowledge, the proposed tests are the only ones available with known asymptotic distribution which can possibly detect such an alternative.

Acknowledgments

Our sincere thanks are due to Professor J. K. Ghosh for his helpful comments. We would also like to thank a referee for suggesting the extension of our original results in terms of Spearman's score function to general score functions.

References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
[2] Bahadur, R. R. (1960). Stochastic comparison comparison of tests. *Ann. Math. Statist.* **31**, 276–295.
[3] Bhuchongkul, S. (1964). A class of nonparametric tests for independence in bivariate populations. *Ann. Math. Statist.* **35** 138–149.
[4] Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32** 485–498.
[5] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric tests statistics. *Ann. Math. Statist.* **29** 972–994.
[6] Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. Wiley, New York.
[7] Loéve, M. (1963). *Probability Theory*. Van Nostrand, Princeton, N.J.
[8] Puri, M. L., Sen, P. K., and Gokhale, D. V. (1970) On a class of rank order tests for independence in multivariate distributions. *Sankyhā Ser. A*. **32** 271–298.
[9] Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

[10] RUYMGAART, F. H., SHORACK, G. R., AND VAN SWET, W. R. (1972). Asymptotic normality of nonparametric tests for independence. *Ann. Math. Statist.* **43** 1122–1135.

[11] SIMON, G. (1977). Multivariate generalization of Kendall's tau with application to data reduction. *J. Amer. Statist. Assoc.* **72** 367–376.

[12] WIEAND, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4** 1003–1011.